PPD pipline

degenerated **sequences**

| | Intron 1 | Exon | Intron 2 |

Sp1, Sp2, Sp3, Sp4, Sp5, Sp6, Sp7, Sp8, Sp9, OG, Target gene

Polymorphic site 1
Polymorphic site 2
Hypervariable sites
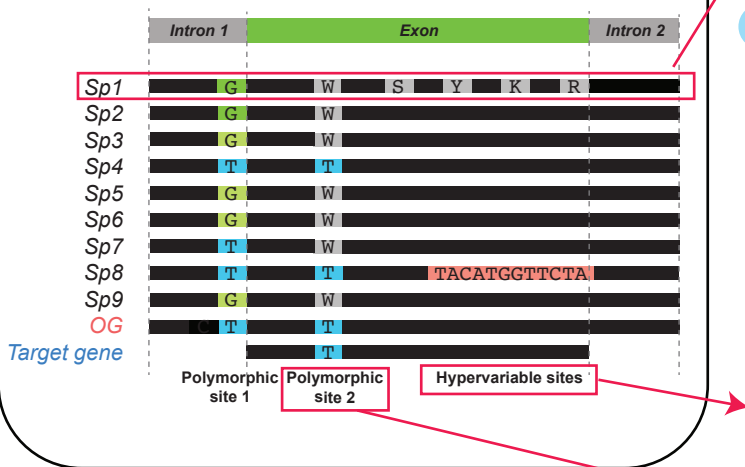
Step 1: resort gene files

Step 2: filter sequences with high percentage of heterozygous sites (0.05%).

Step 3: add target references to gene files.

Step 4: obtain MSA using MAFFT.

Step 5: remove the target references.

Step 6: remove the sites with over 50% gaps using Trimal.

Step 7: remove hypervariable sites using sliding window.

Step 8: Identify putative paralogs using shared heterozygosity in over 50% taxa.